

Williams, D., & Happé, F. (2010). Representing intentions in self and other: Studies of autism and typical development. *Developmental Science*, 13, 307-319.

## Representing intentions in self and other: studies of autism and typical development

David Williams

Institute of Child Health, University College London

Francesca Happé,

Social, Genetic and Developmental Psychiatry Research Centre, Kings College  
London

*This article may not exactly replicate the final version published in Developmental Science.* The published article can be accessed at:

<http://onlinelibrary.wiley.com/doi/10.1111/j.1467-7687.2009.00885.x/full>

## Acknowledgements

This study was supported by a Medical Research Council Studentship, awarded to the first author. The manuscript was prepared during an Economic and Social Research Council Fellowship awarded to the first author.

The authors would like to thank Sophie Lind and Peter Hobson for their helpful comments on an early draft of the manuscript. The authors would also like to thank the staff and students at the schools from which participants were recruited: Cambridge school, Christchurch school, Dalmain school, Garratt Park school, Linden Bridge school, Queensmill School, Springhallow school, St Winifred's school, The Vines school, Turney school.

## Abstract

Two experiments were conducted to explore the extent to which individuals with Autism Spectrum Disorder (ASD), as well as young typically developing (TD) children, are explicitly aware of their own and others' intentions. In Experiment 1, participants with ASD were significantly less likely than age- and ability-matched comparison participants to correctly recognise their own knee-jerk reflex movements as unintentional. Performance on this knee-jerk task was associated with performance on measures of false belief understanding, independent of age and verbal ability, in both participants with ASD and TD children.

In Experiment 2, participants with ASD were significantly less able than comparison participants to correctly recognise their own or another person's mistaken actions as unintended, in a 'Transparent Intentions' task (Russell & Hill, 2001; Russell, Hill & Franco, 2001). Performance on aspects of the Transparent Intentions task was associated with performance on measures of false belief understanding, independent of age and verbal ability, in both participants with ASD and TD children.

This study suggests that individuals with ASD have a limited awareness of their own and others' intentions and that such awareness requires a meta-representational Theory of Mind.

## Introduction

Actions, unlike accidents, are considered to have distinctive mental causes, namely an intention to perform the action and a belief that such an action is possible (e.g., Davidson, 1963). In order to account for actions which are not obviously accompanied by any *conscious* mental state (e.g., changing gear in a car), Searle (1983) has distinguished between two different kinds of intention. A 'prior intention' is a purely mental-representational state which is conscious and which can occur regardless of whether any action is actually performed. An 'intention-in-action', by contrast, serves to guide ongoing action and is often, but not always (as in the example of automatically changing car gears), explicitly conscious.

The upshot of this position is that to grasp the distinction between an action and an accident, a person may need to recognise the mental states – specifically, the intentions – that underlie behaviour. In this context, several studies of young typically developing (TD) infants have indicated the presence of a rudimentary appreciation of others' intentions from at least 18 months of age. In a study by Meltzoff (1995; see also, Bellagamba & Tomasello, 1999), 18-month-olds saw an adult either successfully completing an intended action (e.g., pulling apart two halves of a dumbbell), or trying to complete the action but failing (the adult's hand would 'accidentally' slip off the dumbbell, resulting in the two halves failing to separate). Meltzoff found that when infants in this latter 'failed attempt' condition were offered the materials they had seen the adult act upon, they did not simply copy the surface behaviour they had observed (i.e., slipping), but would actually complete the action the adult had *intended* but failed to achieve (pulling the dumbbell apart). Indeed, infants in the failed attempt condition pulled the dumbbell apart as frequently as infants in the 'full demonstration' condition, who had observed the adult complete the action. This finding suggests that from at least 18 months, infants can accurately distinguish the mental states of the actor from the behaviour the actor actually displays. Similarly, Carpenter, Akhtar, and Tomasello (1998) showed that both 12- and 18-month-olds would copy an adult's action when the action was accompanied by the vocalisation 'There!' (to signal the action had been intentional), but not when the action was accompanied by the vocalisation 'Whoops!' (to signal that the action was unintended). In this case, young infants appear to distinguish very similar actions on the basis of the actor's underlying goal.

Although the results of studies employing imitation paradigms to explore mental state recognition in young infants are not uncontroversial (see Huang, Heyes, & Charman, 2002), they do suggest that some aspects of others' minds are registered by infants from at least 18 months of age. However, it is highly debateable whether young infants/children have the same grasp of intentions as older children do. Several researchers have argued that young children, before approximately 4 years of age, do not properly distinguish intentions from desires (e.g., Astington, 1993; Feinfield et al., 1999; Moses, 1993). Desires and intentions are similar in many ways. Both are mental states which are either fulfilled or unfulfilled, unlike mental states like beliefs which are either true or false, and much of the time desires and intentions coincide in any particular individual. However, the two mental states are not synonymous, as is evident from the fact that one can desire an outcome that one has no intention to bring about and, conversely, one can intend to do something that one does not desire. Astington (1993, pp.91-92) highlights the crucial distinction between desires and intentions as follows, "Desires are fulfilled so long as the outcome is achieved, it doesn't matter how, but intentions are carried out only if the intention causes the

action that achieves the outcome”. Now, the results provided by Meltzoff (1995) and Carpenter et al. (1998), amongst others, do not establish whether infants represent (and therefore copy) the intentions underlying others’ behaviour, or whether their representation is of an undifferentiated desire-intention state. In other words, infants could simply have produced an action which they interpreted as having been *desired* by the adult actor, rather than intended as such.

Some researchers (e.g., Perner, 1991) have suggested that recognition of intentions as mental states that are differentiated from desires emerges later in development and depends upon a meta-representational theory of mind (ToM). Traditionally, performance on so-called false belief tasks is seen as the ‘acid test’ of a meta-representational ToM (Dennett, 1978). Typically developing children pass such tasks from approximately age 4 to 5 years (Wellman, Cross & Watson, 2001) and, importantly, there is evidence that their understanding of intentional action undergoes a developmental shift in this same period. For example, in line with the arguments of Astington (1993) and others, young TD children have difficulty representing intentions (independently of desires), in self or other, in scenarios where the desire is not obvious (e.g., Astington & Lee, 1991; Feinfield, Lee, Flavell, Green & Flavell, 1999) or when the desire conflicts with the intention (e.g., Feinfield et al., 1999; Phillips, Baron-Cohen & Rutter, 1998).

In the study by Phillips et al. (1998), children attempted to hit a pre-specified target in a rigged shooting game. Some of the targets contained prizes whilst others did not. Phillips et al. found that 5-year-olds were able to report their prior intention to hit a particular target whether or not their intention was fulfilled and regardless of whether or not the target yielded a prize (i.e., whether or not their desire was satisfied). On the other hand, the pattern of performance shown by 4-year-olds on this task suggested that they did not understand their own intentions as differentiated from their own desires. When there was a discrepancy between their intention and their desire, 4-year-olds tended to perform poorly. For example, unlike 5-year-olds, if they succeeded in hitting the target of their choice (i.e., fulfilled their intention) but that target did not contain a prize (i.e., did not satisfy their desire), 4-year-olds denied having intended to hit the target.

Further suggesting that young children have difficulty in explicitly representing their own intentions, Shultz et al. (1980) found that children below approximately 5 years of age tended to incorrectly report their own knee-jerk reflex as an intentional action. Reflexes, unlike deliberate actions, are accompanied neither by desires nor intentions. With no possibility of matching goals and outcomes, one may require a coherent concept of intention, differentiated from desire, in order to recognise a reflex movement as an ‘accident’. Further evidence for a relationship between understanding reflexes and a meta-representational ToM was provided by Lang and Perner (2002). They found that the performance of TD children on an unexpected contents false belief task was highly correlated with performance on the knee-jerk task, even after the influences of age and verbal ability were controlled.

Using a different methodology, Russell, Hill and Franco (2001) found further evidence that young TD children do not possess a coherent concept of intention. In a ‘Transparent Intentions’ task, children were asked to complete a drawing of, for instance, a boy with a missing ear, on a transparency like those used with overhead projectors. Unknown to the child, a second transparency with a different, unfinished drawing (e.g., of a cup with a missing handle) was laid in precise alignment on top of the first transparency. Therefore, participants ended up unintentionally finishing off the top drawing (of a handle on a cup) rather than the bottom drawing (of an ear on a

boy). When their mistake was revealed, children were asked both what they had *meant* to draw and what they had *thought* they were drawing throughout their action. These two types of test question were designed to assess participants' awareness of their own prior intentions and intentions-in-action, respectively (see below and also Russell et al., p.780 for further discussion). Russell et al. also included an Other-person condition in which children observed a glove puppet performing the same actions and making the same errors.

Findings from Russell et al.'s (2001) study indicated that 4-year-olds were significantly better than 3-year-olds at answering correctly what they had *meant* to draw. In contrast, 3-year-olds were as able as 4-year-olds to correctly distinguish between what they had *thought* they were drawing and what they had mistakenly drawn. The same pattern also held in the Other-person condition. Finally, Russell et al. found that performance on the Transparent Intentions task (as measured by a composite of performance on the four test questions, across Self and Other-person conditions) was *not* related significantly to ToM (as measured by a composite of performance across three different ToM tests) in these TD participants.

It is debateable, however, whether the different types of Transparent Intentions test question are measuring the same underlying knowledge of intentions. Russell et al. (2001) suggest, first (p.780), that *both* the 'Mean' and 'Think' question-types assess awareness of intentions-in-action. In a footnote (p.780), however, they seem to suggest that the Mean question refers to a prior intention, whereas the Think question refers more to an intention-in-action: "one can regard the mean question as referring to a discrete mental event and the think question as referring to a continuous experience. An intention-in-action, in contrast to a prior intention, would seem to invite the use of continuous aspect". If these different forms of question measure different underlying knowledge, it is questionable why Russell et al. implemented a composite measure of performance across both types of question for comparison with ToM performance. If only one of the question types reliably measures knowledge of the mental-representational aspects of intention, then perhaps answering only these questions (in Self and Other conditions) requires ToM, whereas answering the other type of question does not.

### *Understanding of intentions in Autism Spectrum Disorder*

Autism Spectrum Disorder (ASD) is a condition diagnosed on the basis of severe impairments in social interaction and communication, and a rigid and repetitive behavioural repertoire (American Psychiatric Association, 2000). On the cognitive level, ASD is characterized by deficits in ToM. Individuals with ASD show delayed (e.g., Happé, 1995) and atypical (e.g., Leslie & Thaiss, 1992) performance on false belief tasks, and in everyday life their social and communicative difficulties suggest they do not fully grasp the mental states underlying people's behaviour.

Studies exploring the awareness of intentions by individuals with ASD have produced mixed results. Aldridge, Stone, Sweeney, and Bower (2000) found that young, preverbal children with ASD were unexpectedly competent at completing the 'failed attempts' of actors, in the behavioural re-enactment paradigm implemented by Meltzoff (1995) and Bellagamba and Tomasello (1998) with TD infants. This finding led Aldridge et al. (p.294) to claim that "preverbal children with autism understand the intentions of others". However, as discussed above, such imitation paradigms provide little information about the content of children's representations of mental states. Even if one accepts that successful performance on such tasks indicates a rudimentary awareness of mind (and see Huang et al., 2002, for an ecological, non-

representational account of successful performance), it seems probable that participants need only represent the experimenter's desires, rather than intentions, in order to succeed. In this regard, it is important to highlight how children with autism appear well able to represent the simple desires of other people despite their poor performance on multiple other measures of ToM (Baron-Cohen, 1991; Tan & Harris, 1991).

Together, these findings appear to support the suggestion made above that successful performance on such imitation tasks does not rely on an awareness of the mental-representational aspects of intentions. Indeed, in the small number of studies exploring intention understanding in ASD that have used different paradigms, results tend to indicate impairment in affected children. For example, Phillips et al. (1998) assessed children with ASD on the previously mentioned target-shooting task that they also implemented with young TD children. Like young TD children, participants with ASD (but not matched comparison participants) performed poorly when there was a discrepancy between their intention and their desire. For instance, when participants with ASD mistakenly hit an unspecified target (i.e., did not fulfil their intention) but that target contained a prize (i.e., fulfilled their desire), they tended to incorrectly claim that they had meant to hit the target. Like young TD children, therefore, individuals with ASD appear to confuse their own desires and intentions (although see Russell & Hill, 2001, Experiment 2, for alternative findings).

In contrast, Russell and Hill (2001, Experiment 3) found that, relative to matched comparison participants with moderate learning difficulties (MLD), children with ASD were relatively unimpaired on the 'Transparent Intentions' task implemented with TD participants by Russell et al. (2001). In the Self condition of the task, Russell and Hill found that when the 'Mean' question ('What did you mean to draw?') was asked before the 'Think' question ('What did you think you were drawing?'), children with ASD were significantly impaired relative to comparison children. Only approximately 50% of participants with ASD correctly recognised that they had not intended the actual outcome. However, no other between-group differences reached significance, leading Russell and Hill (p.326) to conclude that "while there was a weak trend in the transparent intentions task for children with autism to perform at a lower level than children with MLD... there was certainly nothing in these data to justify the term 'autistic deficit'".

There are, however, potential concerns about the data and design of Russell and Hill's (2001) experiment which suggest that the results should be interpreted cautiously. Firstly, the use of a repeated-measures design, in which the order of (Self/Other) condition presentation was counter-balanced, resulted in some of Russell and Hill's participants receiving the Other-person condition of the task first. Having already seen the crucial experimental manipulation – that there were two transparencies even though there only appeared to be one – participants who undertook the Other-person condition first would have had some knowledge that they were being 'tricked' throughout their own action in the subsequent Self condition. This may have altered the nature of the task for these participants. Counterbalancing of conditions could well have confounded results, therefore.

A second difficulty with Russell and Hill's (2001) experiment, which may also have affected the results of Russell et al. (2001), was the nature of the stimuli used. In two out of the four drawing combinations used, the missing piece on each transparency was very similar. Hence, in one combination, the participant intended to draw smoke emerging from the chimney of a house (on the bottom transparency) and ended up drawing smoke emerging from the chimney of a boat (on the top

transparency). In another example, the participant intended to draw a face on a girl and ended up drawing a face on a boy. In these cases, children ended up drawing objects so similar to those that they intended to draw, that there may have been no perceived mistake to report.

One further possibility for the lack of group differences in Russell and Hill's (2001) study was the surprisingly poor performance of their comparison group on a standard unexpected contents false belief task. Nearly two thirds of participants from each diagnostic group failed the Other-person test question from this measure. As such, it is perhaps not surprising that both groups found the experimental task difficult, given its potential relation to ToM. Indeed, the numbers of participants in each group failing each of the four test questions on the Transparent Intentions task was closely in line with the numbers failing the false belief task, although Russell and Hill did not analyse the interactions between these tasks.

The two experiments reported in this paper assessed the explicit awareness children with ASD, as well as young TD children, have of their own and others' intentions. In Experiment 1, participants' awareness of their own knee-jerk reflex movements was assessed. Firstly, it was predicted that participants with ASD would incorrectly report their reflex movements as intentional significantly more often than comparison participants with developmental delay (DD). Secondly, it was predicted that performance on the knee-jerk task would be associated significantly with performance on measures of false belief understanding, independently of chronological age (CA) and verbal mental age (VMA). Regarding TD participants, it was predicted that performance on the knee-jerk task would be significantly associated with false belief understanding, independently of CA and VMA (cf. Perner & Lang, 2002).

In Experiment 2, participants undertook a revised version of the Transparent Intentions task, in which the methodological confounds that may have biased results in the studies by Russell and Hill (2001) and Russell et al. (2001) were eliminated. First, it was predicted that participants with ASD would perform significantly less well than DD participants on each of the four test questions, incorporating both the 'Mean' (i.e., 'What did you *mean* to draw?') and 'Think' (i.e., 'What did you *think* you were drawing?') question-types, across both Self and Other conditions. Second, it was predicted that performance on the Transparent Intentions task would be associated significantly with ToM performance, independently of CA and VMA, in all participants.

## Experiment 1

### Method

#### *Participants*

Ethical approval for this research was obtained from the joint South London and Maudsley NHS Trust/Institute of Psychiatry Research Ethics Committee. Twenty-two children with ASD and 22 comparison children participated in Experiment 1, after parents/guardians had given written, informed consent for their children to be included. The participants in the ASD group had received formal diagnoses, by a trained psychiatrist or pediatrician, of autistic disorder ( $n = 20$ ), Asperger's disorder ( $n = 1$ ), or atypical autism/pervasive developmental disorder not otherwise specified (PDD-NOS;  $n = 1$ ) according to established criteria (American Psychiatric

Association, 2000). All participants in this group attended specialist autism schools, which required a diagnosis of autism, Asperger's syndrome or PDD-NOS for entry into the school. The comparison (developmentally delayed; DD) group consisted of 19 children with general learning disability of unknown origin and 3 TD children who were recruited for comparison with those (3) children in the ASD group who achieved IQ scores in the 'average' range (i.e., over 90). Finally, a separate group of 32 TD children also took part.

### *Background Assessments*

Baseline verbal and non-verbal abilities were assessed by an appropriate measure for the developmental level of each participant. The verbal abilities of 16 (out of 22) children with ASD and 17 (out of 22) DD children were assessed by performance on the Vocabulary and Information subtests of the Wechsler Intelligence Scale for Children – Third Edition UK (WISC-III; Wechsler, 1991). The verbal IQ estimate gained from this short form has high reliability (Sattler, 1992). Because the lowest test age-equivalent offered by the WISC-III is 6 years and 2 months, the VMA of any participant who fell below this level on either of the verbal subtests could not be calculated. Under these circumstances, participants were administered the British Picture Vocabulary Scale – Second Edition (BPVS; Dunn et al., 1997), which offers test age-equivalents down to 2 years and 11 months. In this instance, the verbal abilities of six (out of 22) children with ASD and five (out of 22) DD comparison children were assessed with the BPVS. The verbal ability of all TD children was assessed by the BPVS.

The non-verbal ability of all ASD and DD participants was assessed by the Block Design and Picture Completion subtests of the WISC-III. The performance IQ estimate gained from this short form has high reliability (Sattler, 1992). Due to limited child availability, the non-verbal abilities of two participants with ASD and two DD comparison participants were not assessed<sup>1</sup>. The non-verbal ability of TD children was not assessed.

In addition to the 22 participants with ASD for whom matched DD comparison participants were available, a further 8 participants with ASD completed the knee-jerk task but did not have suitable matches and were not therefore included in the between-participant analyses. However, these participants were suitable for inclusion in subsequent within-participant analyses (assessing relationships between tasks), increasing the power of the analyses. The characteristics of each participant group are presented in Table 1. Statistical analyses showed that the ASD and DD groups were well matched on all variables: CA:  $t(42) = -0.55, p = .58, r = .08$ ; VMA:  $t(42) = 0.28, p = .83, r = .04$ ; VIQ:  $t(42) = 0.80, p = .43, r = .12$ ; PIQ:  $t(38) = 0.86, p = .40, r = .14$ .

Table 1 here

Given that some ASD and DD participants received the Wechsler Scales (Wechsler, 1991), whilst others received the BPVS (Dunn et al., 1997), independent *t*-tests were conducted comparing ASD and DD participants from each sub-sample to ensure adequacy of matching in each case, as well as overall. ASD and DD participants who received the Wechsler Scales were well matched on all variables (all *ts* < 1.14, all *ps* > .27), as were ASD and DD participants who received the BPVS (all *ts* < 0.80, all *ps* > .46).



## *Design and procedures*

### *Knee-Jerk Task*

Knee-jerk reflexes were elicited with a Queens reflex hammer, of the type used by medical doctors. Participants sat on top of a table with their legs hanging comfortably over the edge. In order to ensure that the participants were relaxed and reassured, the experimenter explained that the ‘little hammer’ was used for ‘tapping knees’ and that it didn’t hurt because of the rubber disc on the end. The participant was then offered an opportunity to tap the experimenter’s knee which most participants, including those with ASD, seemed to enjoy. On no occasion was a reflex elicited in the experimenter during this preparatory ‘game’. The experimenter then attempted to elicit a knee-jerk reflex in one of the participant’s legs. If, after two attempts, no reflex was elicited, the experimenter moved to the other leg, reassuring the participant, “Now we’ll tap this knee”. Once a reflex had been elicited, the experimenter looked at the participant and said, ‘Look, your leg moved. Did you mean to move you leg?’. This was the only question asked, a correct response being that the participant had not meant to move their leg and an incorrect response being that they had meant to move their leg. Only participants in whom a reflex was elicited were asked the test question.

### *Theory of Mind Assessment*

As measures of ToM, all ASD and DD participants received a traditional unexpected transfer (‘Sally-Anne’) false belief task as well as an unexpected contents task based on the traditional ‘Smarties’ false belief task (Perner, Frith, Leslie & Leekam, 1989). This ‘Plasters’ unexpected contents task involved both Self and Other-person test questions, regarding the participants’ own previous false belief and another’s false belief, respectively.

All TD participants received the Plasters task and 31 (out of 32) also completed the Sally-Anne task.

## *Results*

Overall, 15/22 (68.2%) participants from the ASD group and 20/30 (66.6%) from the extended ASD group passed the knee-jerk task (i.e., correctly reported their reflex movement as unintentional). This compared to 21/22 (95.5%) DD participants and 17/32 (53.1%) TD participants passing the task. The difference in performance between ASD and DD participants was significant,  $\chi^2(1) = 5.50$ , Fisher’s Exact  $p = .04$ ,  $\phi = .35$ .

Figure 1 about here

### *Relation Between Reflex Understanding and ToM*

Table 2 shows the percentage of ASD, DD, and TD participants passing each false belief measure, as well as each group’s mean total score across the three tasks. Participants with ASD (including those from the extended sample) performed significantly less well than DD participants on each measure: Plasters Self:  $\chi^2(1) =$

9.65,  $p = .002$ ,  $\phi = .43$ ; Plasters Other-person:  $\chi^2(1) = 5.30$ ,  $p = .02$ ,  $\phi = .38$ ; Sally-Anne:  $\chi^2(1) = 3.42$ ,  $p$  (one-tailed)  $< .04$ ,  $\phi = .26$ ; Total False Belief Score:  $t(49.91) = -2.91$ ,  $p = .005$ ,  $r = .38$ .

Given the theoretical and empirical links between understanding reflex actions and ToM, a series of Chi-Square analyses were conducted to assess the degree of association in each group between performance on the knee-jerk task and performance on each false belief measure.

Table 2 about here

### *ASD Sample*

In the extended sample of 30 participants with ASD, there was a significant association between performance on the Plasters Self question and performance on the knee-jerk task,  $\chi^2(1) = 9.60$ ,  $p = .002$ ,  $\phi = .57$ . The association between performance on the Plasters Other-person question and performance on the knee-jerk task was also significant,  $\chi^2(1) = 4.34$ ,  $p = .04$ ,  $\phi = .38$ . Finally, there was a significant association between performance on the Sally-Anne task and performance on the knee-jerk task,  $\chi^2(1) = 7.18$ , Fisher's Exact  $p = .007$ ,  $\phi = .49$ .

In this expanded sample of ASD participants, success on the knee-jerk task was significantly related to VMA,  $r_{pb} = .40$ ,  $p = .03$ , and CA,  $r_{pb} = .31$ ,  $p$  (one-tailed)  $= .05$ . The next step, therefore, was to ensure that the above associations between knee-jerk task performance and ToM performance were not simply an artifact of each task's relationship with VMA and CA. In order to assess this, the false belief composite score (see Table 2) was employed as a measure of ToM. This composite was deemed valid after initial analyses revealed that performance on each of the three measures was strongly associated, in this sample ( $\phi = .73 - .76$  between the three test questions).

Bivariate point-biserial correlations revealed that performance on the knee-jerk task was significantly associated with the false belief composite score,  $r_{pb} = .53$ ,  $p = .003$ . When CA and VMA were controlled, in a partial correlation, this association remained significant,  $r_{pb} = .40$ ,  $p = .04$ .

### *Developmentally Delayed Sample*

Given that only one comparison participant failed the knee-jerk test, the relationship between reflex understanding and ToM could not be assessed statistically.

### *Typically Developing Sample*

In TD participants, there was a significant association between performance on the Plasters Self question and performance on the knee-jerk task,  $\chi^2(1) = 7.04$ ,  $p = .008$ ,  $\phi = .47$ . In this sample, however, performance on the knee-jerk task was not significantly associated with performance on either the Plasters Other-person question,  $\chi^2(1) = 0.98$ ,  $p = .76$ ,  $\phi = .06$ , or the Sally-Anne task,  $\chi^2(1) = 2.92$ , Fishers Exact  $p = .15$ ,  $\phi = .31$ .

In these participants, success on the knee-jerk task was significantly related to VMA,  $r_{pb} = .35$ ,  $p = .05$ , but not to CA,  $r_{pb} = .09$ ,  $p = .65$ . The next step, therefore, was to ensure that the association between knee-jerk task performance and

performance on the Plasters Self question was not an artifact of any mediating relationship with VMA. A forced-entry logistic regression was therefore conducted with performance on the knee-jerk task (pass/fail) entered as the dependent variable. In the first instance, Plasters Self performance was entered as a predictor variable in the first block and VMA was included as a second predictor variable in block 2. Table 3 displays the regression statistics for this initial analysis.

Table 3 about here

The inclusion of Plasters Self in block 1 resulted in a significantly improved model,  $\chi^2 = 7.53$ ,  $df = 1$ ,  $p = .006$ , accounting for 28% of the variance in knee-jerk task performance ( $R^2 = .28$ ; see Table 3). In block 1, Plasters Self was a significant predictor of knee-jerk task performance. The inclusion of VMA in block 2 did not significantly improve the model,  $\chi^2 = 0.05$ ,  $df = 1$ ,  $p = .83$ . VMA did not account for any additional variance in knee-jerk task performance over and above Plasters Self ( $R^2 = .28$ ;  $\Delta R^2 = .00$ ). However, after the inclusion of VMA in block 2, Plasters Self, as a predictor of knee-jerk task performance, only approached significance.

In a second regression analysis (see Table 4), Plasters Self and VMA were entered in reverse order. The inclusion of VMA in block 1 resulted in a significantly improved model,  $\chi^2 = 4.19$ ,  $df = 1$ ,  $p = .04$ , accounting for 16% of the variance in knee-jerk task performance ( $R^2 = .16$ ). The significance of VMA as a predictor of knee-jerk task performance in block 1 only approached significance,  $\text{Exp}(B) = 0.54$ ,  $df = 1$ ,  $p = .06$ . Following the inclusion of Plasters Self in block 2, the model accounted for 28% variance ( $R^2 = .28$ ). This improvement in the model approached significance,  $\chi^2 = 3.39$ ,  $df = 1$ ,  $p = .07$ . As such, Plasters Self accounted for an additional 12% of the variance in knee-jerk task performance over and above that accounted for by VMA ( $\Delta R^2 = .12$ ), although statistically this addition only approached significance (see Table 4).

Table 4 about here

In order to check that multicollinearity was not biasing the data from the TD sample, collinearity diagnostics were performed. The variance inflation factor (VIF) associated with each predictor variable was 1.95, well below the value of 10 at which Myers (1990, in Field, 2005, p.175) suggests multicollinearity is likely. Further, the tolerance statistic for each predictor variable was .51, well above the value of .20 at, or below, which Menard (1995, in Field, 2005, p.175) suggests provides cause for concern. Finally, Plasters Self had a condition index of 2.43, whilst VMA had a condition index of 11.74. Whilst this is a difference of 9.31 units, Field (2005, p.261) describes a large and worrying difference as 74.43 units. Given these checks, it is unlikely that multicollinearity was biasing the above regression.

## Discussion

As predicted, participants with ASD were significantly less likely than matched comparison participants with developmental disability to accurately report their reflex movements as unintentional. This suggests that individuals with ASD have a diminished awareness of their own intentional states. Such a suggestion is supported by the strong associations observed between performance on each false belief measure and performance on the knee-jerk task, amongst these participants. Importantly, the

association between ToM and performance on the knee-jerk task was not due to any confounding effects of age or general verbal ability, as the correlation between a false belief composite score and success on the knee-jerk task remained significant after the effects of CA and VMA had been controlled. This finding is compatible with Perner's (1991) suggestion that a meta-representational ToM underlies success on the knee-jerk task.

In TD participants, the relationship between ToM and reflex understanding was less straightforward. The ability to represent one's own prior false belief, as indexed by success on the Plasters Self question, was strongly and significantly associated with success on the knee-jerk task. This association was largely independent of general linguistic ability, with a logistic regression analysis showing that Plasters Self performance explained an additional 12% of the variance in performance on the knee-jerk task over and above VMA alone.

Contrary to predictions, however, performance on the knee-jerk task was not significantly associated with performance on any other of the false belief measures, in TD participants. Particularly relevant was the finding that, contra Perner and Lang (2002), performance on the Sally-Anne task was not correlated significantly with performance on the knee-jerk task. Note, however, that whilst not statistically significant, the strength of the relationship between the two tasks was nonetheless moderate in magnitude ( $\phi = .31$ ). The explanation for the non-significance of this effect may lie in the surprising ease with which TD participants in this study passed the Sally-Anne task, with only 6 (out of 32; see Table 2) participants failing. As such, these ceiling effects prevent an accurate assessment of the relationship between Sally-Anne and knee-jerk task performance in this sample of TD participants.

Therefore, Experiment 1 provided evidence that children with ASD are impaired in their understanding of reflex movements and that this impairment is directly related to ToM deficits. The next experiment in this paper explored the extent to which children with ASD, as well as young TD children, could recognize and report their own mistaken actions on a 'Transparent Intentions' task.

## **Experiment 2**

### **Method**

#### *Participants*

Thirty-four children with ASD, 30 comparison children, and 35 TD children completed the Transparent Intentions task. The (developmentally disabled; DD) comparison group consisted of 25 children with general learning disability and 5 TD children who were recruited for comparison with those children in the ASD group who achieved IQ scores in the 'average' range (i.e., over 90). The verbal ability of 14 (out of 34) children with ASD and 17 (out of 30) comparison children was assessed by performance on the Vocabulary and Information subtests of the Wechsler Intelligence Scale for Children – Third Edition UK (WISC-III; Wechsler, 1991). The verbal ability of the remaining 20 participants with ASD and 13 DD participants was assessed by performance on the BPVS (Dunn et al., 1997). Non-verbal ability was assessed by the Block Design and Picture Completion subtests of the WISC-III. Due to limited child availability, the non-verbal ability of five (out of 34) participants with ASD and six (out of 30) DD comparison participants was not assessed. The characteristics of each participant group are presented in Table 5. Statistical analyses

showed that the ASD and DD groups were well matched for: CA:  $t(62) = -0.07, p = .95, r = 0.01$ ; VMA:  $t(59.66) = 0.25, p = .80, r = 0.03$ ; VIQ:  $t(62) = 0.65, p = .52, r = 0.08$ . However, compared to DD participants, ASD participants had significantly higher PIQs:  $t(51) = 2.29, p = .03, r = 0.31$ .

Table 5 about here

Given that some ASD and DD participants received the WISC-III (Wechsler, 1991), whilst others received the BPVS (Dunn et al., 1997), independent  $t$ -tests were conducted on each sub-sample to ensure adequacy of matching in each case, as well as overall.

In the WISC-III sub-sample, participants with ASD (mean CA = 14.10,  $SD = 1.04$ ) were significantly older than DD participants (mean CA = 12.93,  $SD = 1.42$ ),  $t(29) = 2.56, p = .02, r = .43$ . Participants with ASD (mean VMA = 10.13,  $SD = 2.11$ ) also had significantly higher VMAs than DD participants (mean VMA = 8.51,  $SD = 1.68$ ),  $t(29) = 2.38, p = .02, r = .40$ . No other significant differences were found between groups in the Wechsler sub-sample (all  $ts < 1.74$ , all  $ps > .09$ ).

In the BPVS sub-sample, participants with ASD had significantly higher PIQs (mean PIQ = 75.87,  $SD = 24.74$ ) than DD participants (mean PIQ = 53.29,  $SD = 8.69$ ),  $t(20) = 2.32, p = .03, r = .46$ . No other significant differences were found between groups in the BPVS sub-sample (all  $ts < 0.47$ , all  $ps > .64$ ).

## *Design and Procedures*

### *Transparent Intentions Task*

The experimental task, based on that used by Russell and Hill (2001, Experiment 3), involved 2 sheets of transparent acetate (A and B) each with a different (incomplete) drawing on. When the sheets were placed on top of each other, the drawings became perfectly aligned resulting in only one – the drawing on the bottom sheet – being visible. For example, one pair of pictures involved a teacup with a missing handle (picture A) and a choir boy with a missing ear (picture B). When A was placed on B, only B could be seen. Figure 2 shows the experimental pictures (taken from Russell & Hill, p.324) used.

Figure 2 here

In the Self condition, which was always undertaken first, participants were shown what appeared to be a single transparency (i.e., picture B) and asked to label, verbally, the missing part and then draw it in. After this, the transparencies were separated to reveal the top picture (picture A) which the child had unintentionally completed. Participants were then asked a control question, followed by two test questions (the order of the test questions being counterbalanced across participants):

Control question: What did you draw? (pointing to picture A)

‘Mean’ question: What did you mean to draw? If no spontaneous response was given, the experimenter gave a follow-up question: ‘Did you mean to draw an X or a Y?’. The order of referring to the pictures was counterbalanced.

‘Think’ question: Remember, when you were doing the drawing, what did you think you were drawing? If no spontaneous response was given, the experimenter gave a follow-up question: ‘Did you think you were drawing an X or a Y?’. The order of referring to the pictures was counterbalanced.

The control question was included in this study, unlike in Russell and Hill’s (2001) study, to ensure that participants were not answering the test questions based on a perceived match between their (unfulfilled) intention and reality. That is, it was possible for a child to say that they had ‘intended’ to draw a boy’s ear (when, in fact, they had drawn the handle of a tea cup) because they did not recognise the mismatch between their intention and reality (i.e., they believed that they really *had* drawn a boy’s ear and answered in accordance with this misrepresented reality).

In the Other-person condition, which followed immediately after the Self condition, children were told that the experimenter had ‘played this game with another person’ and had filmed it. Participants were then shown a 30-second clip of the experimenter administering the task to a colleague who pretended to be ignorant of the set-up. The clip ended at the moment when the experimenter separated the transparencies to reveal that the colleague had, in fact, drawn on picture A, and not B as they had intended. Participants were then asked the same three questions as in the Self condition, but with regard to the other person’s intentions.

The Other-person condition in this study differed to the Other-person condition in Russell and Hill’s (2001) study, in that they used a glove puppet to complete the drawings. It was felt that a standardised video of an actor undertaking the task was more naturalistic and ecologically valid, however, leading to the use of a video clip in the current study.

Finally, this study differed from Russell and Hill’s (2001) study in using only 2 sets of pictures, which were counterbalanced across Self and Other conditions (see Figure 2, above). This avoided the potential confound identified with the two additional sets of stimulus pictures, used by Russell and Hill, that depicted two very similar objects (see above discussion). The items within each set of pictures used in this study were clearly very different to one another, making correct and incorrect answers to test questions easier to distinguish.

### *Theory of Mind Assessment*

As measures of ToM, participants received the Sally-Anne unexpected transfer and Plasters unexpected contents tasks. Data from the Plasters task was available for 33/34 ASD participants, 29/30 DD participants and 34/35 TD participants. Data from the Sally-Anne task was available for all ASD and TD participants, but only 33/35 TD participants.

### *Results*

Figure 3 illustrates the percentage of ASD, DD, and TD participants passing each test question from the Transparent Intentions task. In the Self condition, participants with ASD performed significantly less well than DD participants on the ‘Mean’ question,  $\chi^2(1) = 8.81, p = .003, \phi = .37$ , but not the ‘Think’ question,  $\chi^2(1) = 1.42, p = .23, \phi = .15$ . In the Other-person condition, participants with ASD performed significantly less well than DD participants on both the ‘Mean’ question,  $\chi^2(1) = 3.71, p = .05, \phi = .24$ , and the ‘Think’ question,  $\chi^2(1) = 6.43, p = .01, \phi = .32$ .

Figure 3 about here

### *Relation Between ToM and Performance on the Transparent Intentions Task*

The percentage of ASD, DD, and TD participants passing each false belief measure, as well as each group's mean total score across the three tasks, is presented in Table 6. Participants with ASD performed significantly less well than DD participants on each measure: Plasters Self:  $\chi^2(1) = 8.50, p = .004, \phi = .37$ ; Plasters Other-person:  $\chi^2(1) = 4.11, p = .04, \phi = .26$ ; Sally-Anne:  $\chi^2(1) = 3.71, p = .05, \phi = .24$ ; Total False Belief Score:  $t(58.28) = -2.84, p = .006, r = .35$ .

Table 6 about here

In order to assess the relationship between ToM and performance on the Transparent Intentions task, a false belief composite score (0 – 3) was assigned to each participant, on the basis of their performance on each of the false belief measures (see Table 6). This composite was deemed valid for participants with ASD after initial analyses revealed that performance on each of the three measures was strongly associated in this sample ( $\phi = .52 - .69$  between the three test questions). In TD participants, performance on the Plasters Self and Plasters Other-person test questions was strongly associated ( $\phi = .59$ ). However, in this TD sample, performance on the Sally-Anne task was associated only moderately with performance on the Plasters Self question ( $\phi = .34$ ) and weakly with performance on the Plasters Other question ( $\phi = .07$ ). Therefore, it is debatable whether a composite consisting of these three measures is valid amongst these participants. However, for the current purposes it was decided to use this composite score (0 – 3) for statistical analyses because (a) it allowed direct comparison of correlations between ToM and Transparent Intentions task performance amongst ASD and TD participants and; (b) exactly the same patterns of correlation between the Transparent Intentions task and ToM were observed when the ToM composite consisted only of performance across the Plasters Self and Other test questions (0 – 2) (see Table 7).

Table 7 displays the point-biserial correlations between ToM and performance on each test question from the Transparent Intentions task. Given that a relationship between ToM and performance on the Transparent Intentions task was predicted *a priori*, all significance values are reported one-tailed.

#### *ASD Sample*

For participants with ASD, initial analyses revealed significant bivariate point-biserial correlations between ToM composite score and each test question (Mean and Think) in each condition (Self and Other) of the Transparent Intentions task. When CA and VMA were controlled, however, only the correlations between ToM composite score and the Think question in the Self condition and the Think question in the Other-person condition remained significant. The correlations between ToM composite score and the Mean test questions in each condition were not significant, after controlling for CA and VMA.

#### *DD Sample*

Since DD participants performed at near ceiling levels on the Transparent Intentions task (over 90% passed each test question, on average), data from this sample were not analysed.

### *Typically Developing Sample*

For TD participants, initial analyses revealed significant bivariate point-biserial correlations between ToM composite score and the Think test question in the Self condition, as well as the Think test question in the Other-person condition. The bivariate correlation between ToM composite score and the Mean test questions in each condition were not significant in this sample of TD participants. When CA and VMA were controlled, the correlations between ToM composite score and the Think questions in each condition remained significant.

Table 7 about here

### *Relative Difficulty of the Question Types*

In order to assess the relative difficulty of the Transparent Intentions Mean and Think test questions, patterns of within-participant performance were explored using McNemar tests, focusing on those participants who passed only one of the two questions in the Self and Other conditions, respectively. Since DD participants performed at near ceiling levels (over 90% passed each test question, on average), data from this sample were not analysed.

In the Self condition of the Transparent Intentions task, three (9.1%) participants with ASD failed the Think question despite passing the Mean question and two (6.1%) participants showed the opposite pattern of performance. For participants with ASD, therefore, the two test questions did not vary systematically in difficulty,  $\chi^2(1) = 14.06$ , McNemar's  $p > .99$ . In TD participants, the pattern was different: whereas 12 (34.3%) TD children failed the Think question despite passing the Mean question, only two (5.7%) children displayed the opposite pattern of performance. For these participants, therefore, the Think question was significantly more difficult than the Mean question in the Self condition of the Transparent Intentions task,  $\chi^2(1) = 0.70$ , McNemar's  $p = .01$ .

In the Other-person condition of the Transparent Intentions task, five (14.7%) participants with ASD failed the Think question despite passing the Mean question and two (5.9%) participants showed the opposite pattern of performance. For participants with ASD, the two test questions did not vary systematically in difficulty,  $\chi^2(1) = 10.46$ , McNemar's  $p = .45$ . In the TD sample, nine (25.7%) participants failed the Think question despite passing the Mean question, whereas only two (5.7%) children displayed the opposite pattern of performance. This difference was very nearly significant,  $\chi^2(1) = 0.70$ , McNemar's  $p = .06$ .

### Discussion

Relative to matched comparison participants, children with ASD were clearly impaired on the Transparent Intentions task. Over one-third of participants with ASD incorrectly reported that both they and another person had (a) intended a mistaken action and; (b) thought that this (mistake) was the action they were performing throughout. In contrast, comparison participants were generally highly proficient at



distinguishing their own and another's unfulfilled intentions from reality. In this regard, comparison participants were significantly superior to participants with ASD, except on the Self 'Think' question which approximately 20% failed. These findings stand in contrast to those of Russell and Hill (2001) who found children with ASD to be impaired, relative to matched comparison participants, on the 'Mean' question in the Self condition only.

Amongst both participants with ASD and TD children, ToM was associated significantly with performance only on the Think questions (in Self and Other conditions) of the Transparent Intentions task, independent of age and verbal ability. This contrasts with Russell et al.'s (2001) finding that ToM was not related to overall performance on the Transparent Intentions task in TD children.

### General Discussion

The findings from the two experiments reported here provide further evidence that individuals with ASD are impaired in their understanding of their own and others' intentions. Relative to age- and ability-matched DD comparison participants, participants with ASD were less able to recognise their own reflex movements as unintentional (in Experiment 1) or their own mistaken actions as unintended (in Experiment 2). In each case, these impairments were associated with a specific difficulty in representing false beliefs – a proxy for ToM – independent of age and verbal ability. In TD participants, also, the ability to attribute false beliefs to self and other was significantly associated with performance on aspects of each of the tasks from Experiments 1 and 2. These findings support the notion that recognising intentions and false beliefs depends upon the same representational mechanism (e.g., Perner, 1991).

It is intriguing that amongst both ASD and TD participants, success on the knee-jerk task in Experiment 1 was most strongly related to success on the Plasters Self false belief question. Note that both semantically and pragmatically, the test questions on the respective tasks are quite different: 'Did you mean to move your leg?' in the knee-jerk task, versus 'What did you think was inside the box before you looked?' in the Plasters Self task. The answers required are also quite different: a yes/no response in the knee-jerk task, versus an open-ended response in the Plasters task. Again, these findings suggest that the underlying representational demands are similar in each task, both requiring, it is argued, recognition of mental states as causal of actions: in the knee-jerk task, there was no intention to move and therefore the movement could not have been 'meant'. In the Plasters Self task, one's (false) belief, rather than reality itself, caused one to choose the plasters box.

The findings from Experiment 2 that participants with ASD were significantly impaired on a Transparent Intentions task clearly differs from the findings of Russell and Hill (2001) which provided little support for an ASD-specific impairment on this task. However, concerns over aspects of Russell and Hill's design, most notably their counterbalancing of the order of presentation of the Self and Other conditions and their use of potentially insensitive stimulus materials, suggest that the present results may be a more accurate reflection of the ability of children with ASD to represent their own (unfulfilled) intentions.

Perhaps the discrepancy between the findings of the current study and those of Russell and Hill (2001) can be explained in terms of the comparison participants used in each study. Comparison participants in Russell and Hill's study performed quite poorly on standard false belief tasks and their performance on the Transparent

Intentions task appeared to reflect this. However, comparison participants in the current study performed significantly better than participants with ASD on a false belief composite measure, which may explain their superior performance on the experimental task. This explanation is partially supported by the finding that amongst both participants with ASD and TD children in the current study, ToM ability was significantly associated with performance on the Think test questions of the Transparent Intentions task, after the effects of age and verbal ability were controlled.

The finding that ToM was related only to performance on the Think questions (in Self and Other-person conditions) of the Transparent Intentions task, but not to the Mean questions (in either condition), suggests that only the former question-type taps a meta-representational ToM. Indeed, for TD participants, the Mean test questions were significantly less challenging than the Think test questions. With hindsight, this result might have been predicted. A number of studies have shown that before they are able to pass false belief tasks, young TD children can report an action as ‘mistaken’ when a behaviourally specified goal does not match an outcome (e.g., Feinfield et al., 1999; Phillips et al. 1998; Shultz et al., 1980). It was argued above that young children’s success in reporting unintended actions in such tasks is based on their recognition of desires, rather than intentions. On this basis, young TD children in Experiment 2 may have succeeded on the Mean test questions of the Transparent Intentions task by defaulting to their concepts of desire, in the absence of a coherent concept of intention. Therefore, this may be cause for concern over the validity of the Transparent Intentions task as a measure of intention understanding in TD children. However, in Experiment 1 TD children had a great deal of difficulty correctly reporting their reflex movements as unintentional, with participants’ performance on the knee-jerk closely associated with their ability to recognise their own false beliefs. The poorer performance of TD children on the test question from the knee-jerk task (‘Did you *mean* to move your leg?’) than the Mean test question from the Transparent Intentions task (‘Did you *mean* to draw X?’) may be explained by the fact that no desire was evident in the knee-jerk task. Hence, unlike on the Transparent Intentions task, participants in Experiment 1 could not succeed on the test question by reporting on their desire, rather than intention, to have acted. Overall, therefore, confidence can still be maintained in the conclusion that young TD children, as well as individuals with ASD, have difficulty in recognising intentions as mental states differentiated from desires.

#### *Wider theoretical implications of the current research*

Several theoretical implications follow from the current set of findings. Specifically in relation ASD, the findings indicate that impairments in representing mental states in self are at least as profound as impairments in representing the mental states of other people. This adds to the growing body of literature indicating diminution of at least some aspects of self-awareness in ASD (e.g., Ben-Shalom et al., 2003; Hobson et al., 2006; but see Williams & Happé, in press a), and in particular Theory of *own* Mind (Frith & Happé, 1999; Williams & Happé, in press b).

More generally, the results provide a direct challenge to models of mindreading that postulate distinct mechanisms for representing one’s own and others’ mental states (e.g., Nichols & Stich, 2003; Raffman, 1999). Nichols and Stich, for instance, claim that there exists one cognitive mechanism for the representation of one’s own propositional attitudes and another mechanism for representing others’ propositional attitudes. Furthermore, they claim that ASD

involves a deficit only in reading the minds of other people, leaving introspection of own mental states intact. The current findings argue strongly against this position, suggesting that (a) individuals with ASD have as much difficulty recognising their own mental states as they do recognising others' mental states and; (b) that one cognitive mechanism (or process) is involved in both recognising one's own and others' mental states (cf. Carruthers, in press). Whether this mechanism/process involves a form of *simulation*, such that knowledge of one's own mental states is employed to read others' minds, or whether introspection of own mental states involves the same *theoretical* underpinnings as other-person mentalising is a matter of debate. We believe there are persuasive arguments why simulation from one's case cannot alone account for the ability to recognise mental states in others (e.g., Carruthers, in press; Hobson, 1990; Strawson, 1962; Williams & Happe, in press b; Wittgenstein, 1953), although it is beyond the scope of this paper to explore these arguments directly. What is clear from the current findings is that, in addition to difficulties with representing false beliefs, individuals with ASD also show deficits in recognising their own and others' intentions.

## References

- Aldridge, M. A., Stone, K. R., Sweeney, M. H., & Bower, T. G. R. (2000). Preverbal children with autism understand the intentions of others. *Developmental Science*, 3, 294-291.
- American Psychiatric Association (2000). *Diagnostic and Statistical Manual: 4<sup>th</sup> Edition Test Revised (DSM IV-TR)*. New York: Author.
- Astington, J.W. (1993). *The child's discovery of mind*. Fontana Press: London, UK.
- Astington, J.W. & Lee, E. (1991, April). *What do children know about intentional causation*. Paper presented at the Biennial Meeting of the Society for Research in Child Development, Seattle, USA.
- Baron-Cohen, S. (1991). The development of a theory of mind in autism: deviance and delay. *Psychiatric Clinics of North America*, 14(1), 33-51.
- Bellagamba, F., & Tomasello, M. (1999). Re-enacting intended acts: comparing 12- and 18-month-olds. *Infant Behavior and Development*, 22, 277-282.
- Ben-Shalom, D., Mostofsky, S.H., Hazlett, R.L., Goldberg, M.C., Mcloud, D., & Hoehn-Saric, R.L. (2003). Intact GSRs and impaired self-reports in response to emotional pictures in high-functioning autism. *Annals of the New York Academy of Sciences*, 985, 501-504.
- Carpenter, M., Call, J., & Tomasello, M. (2005). Twelve- and 18-month-olds copy actions in terms of goals. *Developmental Science*, 8(1), F13-F20.
- Carruthers, P. (in press). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*.
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23), 685-700.
- Dennett, D.C. (1978). *Brainstorms*. Montgomery, VT: Bradford Books.
- Dunn, L.M., Dunn, L.M., Whetton, C., & Burley, L. (1997). *British Picture Vocabulary Scale (2<sup>nd</sup> Ed.)*. NFER-Nelson.
- Feinfield, K.A., Lee, P.P., Flavell, E.R., Green, F.L., & Flavell, J.H. (1999). Young children's understanding of intention. *Cognitive Development*, 14, 463-486.
- Field, A. (2005). *Discovering statistics using SPSS (2<sup>nd</sup> Ed.)*. London, UK: Sage.
- Frith, U., & Happe, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language*, 14, 1-22.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66, 843-855.

- Hobson, R. P. (1990). On the origins of self and the case of autism. *Development and Psychopathology*, 2, 163-181.
- Hobson, R. P., Chidambi, G., Lee, A., & Meyer, J. (2006). Foundations for self-awareness: an exploration through autism. *Monographs of the Society for Research in Child Development*, 71.
- Huang, C-T, Heyes, C. M. & Charman, T. (2002). Infants' behavioral re-enactment of 'failed attempts': exploring the roles of emulation learning, stimulus enhancement and understanding of intentions. *Developmental Psychology*, 38, 840-855
- Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology*, 20, 67-76.
- Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition*, 43, 225-251.
- Meltzoff, A.N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental Psychology*, 31, 1-16.
- Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.
- Moses, L.J. (1993). Young children's understanding of belief constraints on intention. *Cognitive Development*, 8, 1-25.
- Myers, R. (1990). *Classical and modern regression with applications*. Boston, MA: Duxbury.
- Nichols, S., & Stich, S. (2003). *Mindreading: an integrated account of pretence, self-awareness, and understanding others minds*. Oxford University Press.
- Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.
- Philips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and autism. *British Journal of Developmental Psychology*, 16, 337-348.
- Raffman, D. (1999). What autism may tell us about self-awareness: a commentary on Frith and Happe. *Mind & Language*, 14(1), 23-31.
- Russell, J., & Hill, E. L. (2001). Action monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry*, 42, 317-328.

- Russell, J., Hill, E., & Franco, F. (2001). The role of belief veracity in understanding intentions-in-action: preschool children's performance on the transparent intentions task. *Cognitive Development*, 16, 775-792.
- Sattler, J.M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement*. USA: Jerome M. Sattler Publishers Inc.
- Searle, J. (1983). *Intentionality*. Cambridge: Cambridge University Press.
- Shultz, T. R., Wells, D., & Sarda, M. (1980). Development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology*, 19, 301-310.
- Strawson, P. F. (1962, originally 1958). Persons. In V. C. Chappell (Ed.), *The philosophy of mind* (pp.127-146). Englewood Cliffs, NJ: Prentice-Hall.
- Tan, J., & Harris, P. L. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology*, 3, 163-174.
- Wechsler, D. (1991). *Wechsler intelligence scale for children*. (Third Edition), London, UK: The Psychological Corporation.
- Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development*, 72, 655 – 684.
- Williams, D., & Happé, F. (in press a). Pre-conceptual aspects of self-awareness in autism spectrum disorder: the case of action monitoring. *Journal of Autism and Developmental Disorders*.
- Williams, D., & Happé, F. (in press b). ‘What did I say?’ versus ‘what did I think?’: attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders*.
- Wittgenstein, L. (1953). *The philosophical investigations*. New York: The Macmillan Company.

**Table 1: Participant characteristics for Experiment 1: Means and (standard deviations)**

	ASD	ASD-Extended	DD	TD
<i>n</i>	22	30	22	32
CA: yrs	10.61 (3.51)	10.02 (3.21)	11.20 (3.56)	4.43 (0.68)
VMA:yrs	7.69 (3.06)	7.02 (2.90)	7.51 (2.42)	4.86 (1.24)
VIQ	77.32 (17.38)	76.50 (15.58)	73.32 (15.71)	104.1 (10.40)
PIQ	77.95 (22.06) <sup>a</sup>	78.88 (20.50) <sup>b</sup>	72.10 (21.19) <sup>a</sup>	—

<sup>a</sup> Due to limited child availability, PIQ data was available for  $n = 20$  participants with ASD and  $n = 20$  comparison participants.

<sup>b</sup> In the Extended ASD sample, PIQ data was available for  $n = 24$  participants

***Table 2: Percentage of participants from each group in Experiment 1 passing the Plasters Self, Plasters Other and Sally-Anne tasks, as well as overall performance across the three tasks***

	ASD-extended ( <i>n</i> = 30)	DD ( <i>n</i> = 22)	TD ( <i>n</i> = 32)
Plasters Self	50.0	90.9	37.5
Plasters Other	56.7	86.4	56.3
Sally-Anne	63.3	86.4	83.9*
Total False Belief Score (out of 3)	1.70 ( <i>SD</i> 1.37)	2.64 ( <i>SD</i> 0.95)	1.81 ( <i>SD</i> 1.04)*

\*Based on 31/32 TD participants



**Table 3: Summary of logistic regression analysis for variables predicting knee-jerk task performance**

Variable	<i>B</i> ( <i>SE</i> )	Wald	<i>p</i>	95% confidence interval for exp <i>B</i>		
				Lower	exp <i>B</i>	Upper
Block 0						
Constant	-0.13 (0.35)	0.13	.72		0.88	
Block 1 <sup>a</sup>						
Constant	-1.61 (0.78)	4.32	.04		0.20	
Plasters Self	2.29 (0.91)	6.06	.01	1.58	9.29	54.77
Block 2 <sup>b</sup>						
Constant	-1.01 (2.78)	0.13	.72		0.36	
Plasters Self	2.05 (1.20)	2.95	.09	0.78	7.78	80.91
VMA	-0.10 (0.45)	0.49	.82	0.38	0.91	2.19

<sup>a</sup>At Block 1,  $R^2 = .28$  (Nagelkerke). Model  $\chi^2(1, N = 32) = 7.53, p = .006$ .

<sup>b</sup>At Block 2,  $R^2 = .28$  (Nagelkerke). Model  $\chi^2(2, N = 32) = 7.57, p = .02$ .

**Table 4: Summary of logistic regression analysis for variables (in reverse order) predicting knee-jerk task performance**

Variable	B (SE)	Wald	p	95% confidence interval for expB		
				Lower	exp B	Upper
Block 0						
Constant	-0.13 (0.35)	0.13	.72		0.88	
Block 1 <sup>a</sup>						
Constant	2.88 (1.61)	3.19	.07		17.80	
VMA	-0.63 (0.33)	3.55	.06	0.28	0.54	1.03
Block 2 <sup>b</sup>						
Constant	-1.01 (2.78)	0.13	.72		0.36	
VMA	-0.10 (0.45)	0.49	.82	0.38	0.91	2.19
Plasters Self	2.05 (1.20)	2.95	.09	0.78	7.78	80.91

<sup>a</sup>At Block 1,  $R^2 = .16$  (Nagelkerke). Model  $\chi^2(1, N = 32) = 4.19, p = .006$ .

<sup>b</sup>At Block 2,  $R^2 = .28$  (Nagelkerke). Model  $\chi^2(2, N = 32) = 7.57, p = .02$ .

**Table 5: Participant characteristics for Experiment 2: Means and (standard deviations)**

	ASD	DD	TD
<i>n</i>	34	30	35
CA: years	11.06 (3.59)	11.12 (3.55)	4.46 (0.67)
VMA: years	7.48 (2.79)	7.33 (2.00)	4.81 (1.23)
VIQ	76.03 (17.70)	73.10 (18.60)	103.6 (9.79)
PIQ <sup>a</sup>	75.69 (19.93)	63.67 (17.81)	–

<sup>a</sup> Due to limited child availability, PIQ data was collected for  $n = 29$  participants with ASD and  $n = 24$  DD participants.

**Table 6: Percentage of participants from each group in Experiment 2 passing the Plasters Self, Plasters Other and Sally-Anne tasks, as well as overall performance across the three tasks**

	ASD ( <i>n</i> = 34)	DD ( <i>n</i> = 30)	TD ( <i>n</i> = 35)
Plasters Self	51.5*	86.2*	37.1*
Plasters Other	63.6*	86.2*	45.7*
Sally-Anne	70.6	90.0	82.9*
Total False Belief Score (out of 3)	1.85 ( <i>SD</i> 1.23)*	2.63 ( <i>SD</i> 0.90)*	1.73 ( <i>SD</i> 1.04)**

\*Based on 33/34 ASD participants, 29/30 DD participants, and 34/35 TD participants

\*\*Based on 33/35 TD participants

**Table 7: Correlations between ToM total score (0 – 3) and performance on each test question from the Transparent Intentions task. Figures in brackets are partial correlations after controlling age and verbal ability**

Transparent Intentions test question	ToM total score	
	ASD	TD
Self: Mean	.57*** (.20)	.08
Self: Think	.78*** (.61***)	.63*** (.51**)
Other-person: Mean	.48** (.16)	.28
Other-person: Think	.59*** (.34*)	.50** (.43**)

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .05$

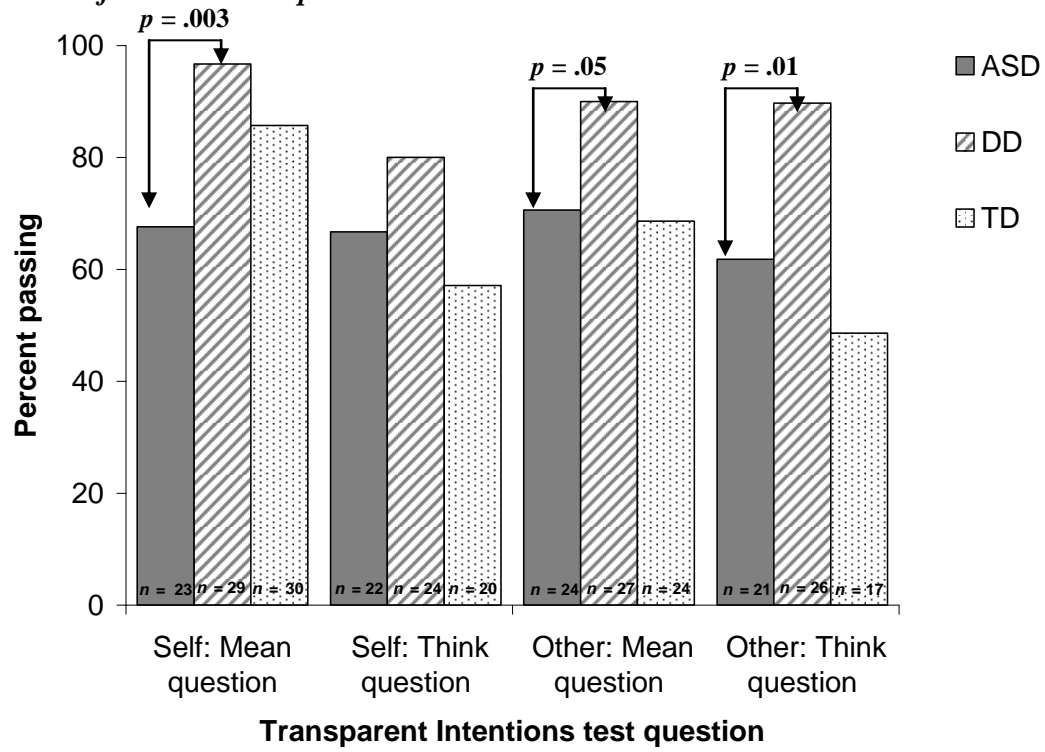
***Figure 1: Experimental materials for the Transparent Intentions task***

**Picture B (bottom transparency)**

**Picture A (top transparency)**



**Figure 2: Percentage (and n) of ASD, DD and TD participants passing each test question from the Transparent Intentions task**



#### Footnotes

1. Given that performance IQ data was not available for two ASD and two comparison participants we wanted to rule out the possibility that our groups could have become significantly different from each other on these ability measures if these individuals had been assessed. To ensure this, we arbitrarily assigned the two outstanding ASD participants with the minimum performance IQ score possible (45 points) and the two remaining comparison participants with the maximum performance IQ score (140 points). We then re-analysed the data. Group comparisons on these measures yielded the following results:  $t(42) = -0.43$ ,  $p = .67$ ,  $r = .07$ . The effect size for this comparison was small (Cohen, 1992) and, thus, both groups would still have been well matched even if these two comparison participants had achieved extremely high performance IQ scores whilst the two ASD participants had achieved extremely low performance IQ scores.